CSCE 5290: Natural Language Processing

FRAUD DETECTION BY USING TEXT CLASSIFICATION

Project Proposal Description:

1a) Project Title:

We are decided to propose the project of Using the **Fraud Detection by Using the Technique of Text Classification** based on the techniques of Natural Language Processing like Text classification and Named entity Recognition.

1b) **Team Members:**

Group-07

1.Venkata Manisahith

2.Lokesh Naidu Bavigadda

3.Srikanth Karni

4.Pasumarthi Raghu Ram

GitHub Link: Link

Goals and Objectives:

- Motivation:
  - A project employing text categorization for fraud detection may have the goal of increasing the effectiveness and efficiency of

fraud detection systems. Fraudulent actions have the potential to harm both enterprises and people financially. Organizations can more effectively spot and stop fraudulent behavior before it causes harm by employing text classification.

- The workload of human analysts who would typically need to manually go through a lot of emails and transactions to find fraudulent behavior can be lessened with text classification. Organizations may automate the process of detecting fraud with the aid of machine learning models, allowing them to react to possible risks more rapidly.

- Using NLP and ML techniques to classify text data as fake or genuine offers a more efficient and accurate solution to fraud detection. The project aims to analyze and process large volumes of textual data such as customer reviews, emails, and transaction descriptions, which can be challenging for traditional methods.

• **Significance:**

1. A text classification-based fraud detection project stands out from other projects in the market due to its ability to accurately detect fraudulent emails, transactions, and messages by analyzing their content.
2. This project offers several advantages over competing projects, such as customization to fit specific business needs, real-time detection to enable quick intervention, and the use of hybrid models that combine various machine learning techniques and data sources to create a more powerful system.
3. Additionally, the results produced by this project can be explained, leading to increased transparency and trust in the system.

4. Finally, the project's ability to adapt to new fraud types and continuously improve its performance using active learning and transfer learning strategies makes it a valuable tool for businesses seeking to identify and prevent fraudulent activities.

In summary, text classification-based fraud detection projects provide unique benefits that make them an important choice for businesses looking to detect and prevent fraudulent behavior.

- **Objectives:**
  - By considering the common problems we decided to make the project as its objectives as
  - To save time and resources: Traditional fraud detection techniques can be time consuming and use many resources which can be expensive process with addition they loss the customer which means brand damage. Here our project automates the process very fast in real time.
  - To reduce Fraudulent activities in number: Due its accuracy and fastness it can be able to reduce the activities by fraudsters by using this algorithm.

- **Features:**
  - Data collection module: collecting samples of both legitimate and fraudulent messages or transactions in a big and varied dataset of text data.
  - Data pre-processing module: deleting stop words, punctuation, and other extraneous information from the text, cleaning it up, putting it in a standard format.
  - Feature extraction module: utilizing methods like bag-of-words, n-grams, and word embeddings to extract pertinent features from the pre-processed text data.
  - Model selection and training module: Selecting a suitable machine learning algorithm and training it on the pre-processed and feature-

extracted data. Examples of such algorithms include Naive Bayes, SVM, and deep learning models like LSTM.

- Model evaluation module: assessing the trained model's performance on a test dataset using metrics like recall, accuracy, and precision.
- Real-time detection module: creating a project with low latency, high throughput, and real-time fraud detection.
- Customization module: adjusting the project to match the unique requirements of a certain sector, group, or use case.

References:

Here are some of the references that can be useful for a text classification project for fraud detection:

- Zhang, X., Li, Q., Huang, X., & Wu, X. (2018). A deep learning approach for financial fraud detection based on textual data. Expert Systems with Applications, 107, 12-21.
- Abbasi, S. A., & Chen, H. (2008). Cybercrime classification: A motivational model. Communications of the ACM, 51(3), 88-93
- Wang, Z., Sun, Y., & Gao, X. (2020). A novel credit card fraud detection model based on word embedding and deep learning. Expert Systems with Applications, 156, 113397.
- Kumar, N., Agarwal, A., & Choudhary, A. (2020). Fraud Detection in E-commerce: A Comprehensive Review. Journal of Ambient Intelligence and Humanized Computing, 11(10), 4235-4261